Comparing Traditional Maps with Twitter-Derived Maps: Exploring Differences and Similarities

INPUT 2016

9th International Conference
on Innovation in Urban
and Regional Planning

input TORINO 2016

e-agorà | e-αγορά

for the transition toward resilient communities

edited by G. Colombo | P. Lombardi | G. Mondini

9th International Conference on Innovation in Urban and Regional Planning

e-agorà/e-ἀγορά for the transition toward resilient communities

Conference Proceedings Book

ISBN 978-88-9052-964-1

INPUT 2016 "e-agorà/e-άγορά for the transition toward resilient communities"

Conference Proceedings from the INPUT2016 Conference in Turin (14th–15th September 2016)

Edited by Giovanni Colombo (ISMB Istituto Superiore Mario Boella), Patrizia Lombardi (DIST - Politecnico di Torino), Giulio Mondini (SiTI Higher Institute on Territorial Systems for Innovation)

Editorial coordination by Stefania Mauro

Graphic design by Sara Oggero (ISMB)

ISBN 978-88-9052-964-1

Table of Content

INPUT 2016 is the ninth meeting with the name "INPUT"

A biennial appointment that started in 1999 in Venice at the IUAV.

We had two shifts in the conference, one in date: 2005 (Alghero) was followed by 2008 (Lecco), to avoid overlapping with the CUPUM conference (*Computers in Urban Planning and Urban Mangament*); and one in name: the acronym INPUT at the beginning stood for *INformatica e Pianificazione Urbana e Territoriale* and now it's *Innovazione e Pianificazione Urbana e Territoriale*.

I have been one of the organizer of the first meeting and I'm very proud of the results the initial intuition has yielded through the organization of this long series of conferences.

In 9 conferences all across the country (Venezia, Isole Tremiti, Pisa, Alghero, Lecco, Potenza Cagliari, Napoli, and now Torino) hundreds of experts and users had the opportunity to share ideas, experiences, tools and projects; people from academic world (among them: urban planners, architects, engineers, computer scientists, sociologists), public administration, and industry (from small start-ups to big enterprises) have had the opportunity to explore and measure the relevance of the ICT for the new ways to think and practice planning and design.

Now we have to face new challenges and maybe rethink the formula of the conference.

We know we were right because nowadays one of the most common sense and mainstream expression is "smart city" (personally I'm not fond of this expression, but it is a fact that this is an expression widely used); we know we have to change exactly for the same reason: we need to avoid the abuse of that expression that can lead to an overly technocratic approach often imbued with ideology; as usual we need to make use of the best available technologies, but having an idea of the purpose of planning, a shared vision of the future.

For this reason I am wondering if this occasion could be the moment for a step forward: from the birth of an Association, to the organisation of a seminar for young researchers and professionals (one year the biennial conference, the next year the seminar), to the opening of the conference to other disciplines (history, restoration, archaeology, …).

The Torino conference could be the right occasion for this *shift of perspective*: among its organisers, in addition to the *Interuniversity Department of Regional and Urban Studies and Planning of the Politecnico di Torino and Università di Torino* (DIST), there are two research institutes: *Istituto Superiore on Territorial Systems for Innovation* (SiTI) and *Istituto Superiore Mario Boella on the Information and Communication Technologies* (ISBM); so that research, education, applications and projects are all brought together through the experiences of the organizing institutions: a good viaticum for the future course of INPUT.

Arnaldo Cecchini

INPUT 2016 "e-agorà/e-άγορά for the transition toward resilient communities"

It is universally recognised that the *Smart City* perspective raises a wide spectrum of unexplored and interdependent problems and extends the horizon over which the *City* growth strategies are defined. Energy generation and consumption models, urban mobility schemes, service processes, goods production mechanisms, citizens' behaviour and community habits are all aspects radically challenged by this perspective. These are sufficient circumstances to affirm that the *smart and sustainable* perspective of our cities is fully inscribed in the fundamental questions of our age.
And it is exactly the character of these *fundamental questions* that makes *Smart City* an unrepeatable occasion for society to challenge on subjects of *technical*, *economical*, *territorial* and *societal* nature that need to be stimulated jointly if the essential aim of Smart City is really the *good life* for society. In this framework, it is fundamental that the technical discontinuities are *responsive* ahead of the unprecedented needs of a sustainable development and the financial system is *flexible* enough to support the new kinds of infrastructural solutions. The territorial and urban disciplines are singled out to elaborate *innovative* concepts enabling the completely renewed City processes to take place. The public administration systems must guarantee *effective* measures and incentives to facilitate the inevitable transformations. The societal bodies must play an essential role in increasing the level of *consciousness* and *participation* of the citizens in defining and verifying the suitability of the new social processes.

All these aspects are covered in our Input Conference, where a wide spectrum of scientific thoughts and sensibilities are brought together with the aim of creating a common and challenging perspective: an intelligent, sustainable and inclusive City as a fundamental contribution to the environmental health and the social wellbeing.

Giovanni Colombo

Comparing Traditional Maps with Twitter-Derived Maps: Exploring Differences and Similarities

Stefano Pensa[a] and Elena Masala[b]

[a] DAD - Department of Architecture and Design, Politecnico di Torino, Torino, Italy (stefano.pensa@polito.it)
[b] SiTI - Higher Insitute on Territorial Systems for Innovation, Torino, Italy (elena.masala@polito.it)

Keywords: social media; twitter; urban planning; GIS; visual analytics.

Introduction

Nowadays, data generated by the users of ICT social media are easily available raw data, which often include a geo-information and provide lots of information on individual and collective life. This huge quantity of information is expected to offer opportunity for innovation in urban and transport planning (Bawa-Cavia 2010; Chua, Marcheggiani, Servillo and Vande Moere 2014; García-Palomares, Gutiérrez and Mínguez 2015; Hahmann, Purves and Burghardt 2014; Kokalitcheva 2014; Lanzerotti, Bradach, Sud and Barmeier 2013; Kwan 2016; Neuhaus 2011; Zachariadis *et al.* 2015). Nevertheless, the application in the planning practice presents some points of discussion. Miller and Goodchild (2015) outline how data-driven geography should consider the necessity of filtering user-generated data because of their messiness and the need for integration with other information because data have no answer for everything. In particular, they remarks that social media data are not a sample of the whole population but the whole set of data produced by a self-selected population. In order to understand how social media data can be used to integrate, or even to substitute, traditional data, this paper describes the outcomes from a comparison between traditional official maps and maps derived from the analysis of Twitter data in Pampulha, a neighbourhood of Belo Horizonte, Brazil. The comparison focuses on land uses and transport infrastructures, considered as the basic elements for the elaboration of an urban plan.

Methodology

Available data for the case study

Pampulha is an administrative region of Belo Horizonte, Brazil, grown on the surrounding of a man-made lake built in the early 1940s. Pampulha is a neighbourhood particularly attractive for both its citizens and tourists, due to the architectures of Oscar Niemeyer, gardens by Burle Marx and paintings by Candido Portinari, but also for hosting the Universidade Federal de Minas Gerais (UFMG), the "Mineirão" Stadium, the Mineirinho Arena, the Zoo, a small airport, industrial zones, different urban gardens and a military area. This mix is resumed in the official land use map (Figure 1).

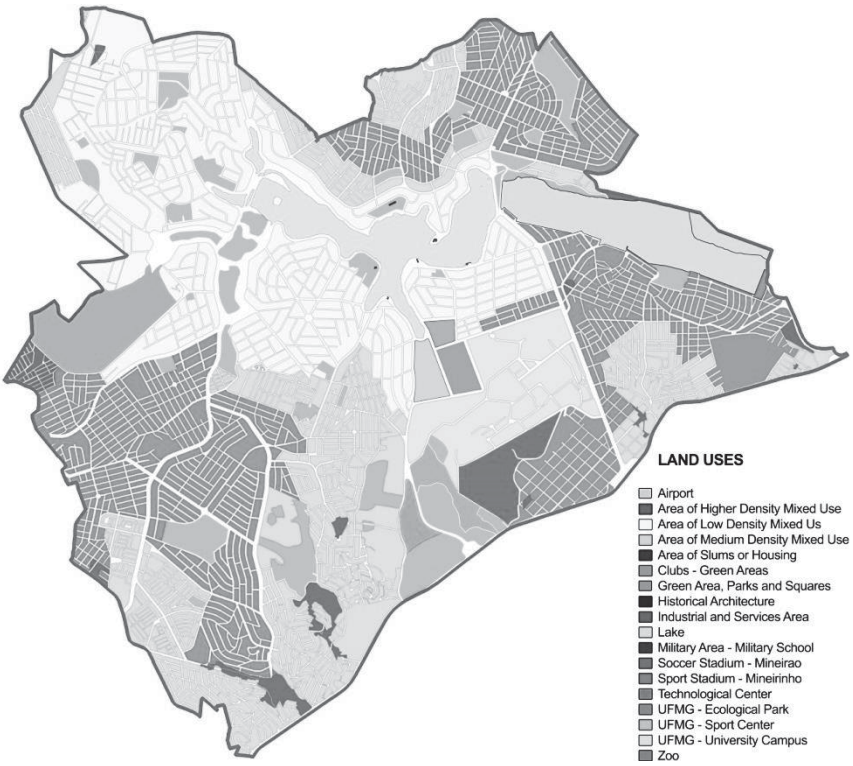

Fig. 1. Official map of land use in the area of Pampulha, Belo Horizonte, Brazil. Source: Prodabel/PBH - Municipality of Belo Horizonte, 2015.

With regards to the transport planning of the city, the official map provides information on the capacity of roads expressed in term of number of vehicles which can pass in one hour. In order to make a comparative analysis, this research makes use of Twitter data, collected between March 26th and July 22nd, 2015, from the whole metropolitan area of Belo Horizonte. Collected data are only geo-referred tweets, whose amount is 929,281 tweets sent by a total of 41,317 different users.

Methodology

The first step consisted in filtering raw data in order to obtain a reduced database containing only the tweets of users who sent almost one tweet from Pampulha neighbourhood. All following analysis were performed on this filtered database, which resulted in 42,991 tweets sent by 5,173 different users.

A second step consisted in calculating the density of tweets, which was processed by applying the quartic kernel function (Silverman 1986, p. 76, equation 4.5). Through this function, sent tweets

were grouped in areas with the same level of density. The same function was applied also to the official land-use map, providing a map showing a level of density in both the cases: density of land use for the traditional map, and density of Twitter activity for the tweet footprint map (Figure 2).

The third step consisted in elaborating a map containing the possible movements of Twitter users. Thus, data were elaborated to correlate the subsequent tweets of each single user. Physical and temporal distances between each couple of tweets were calculated in order to select only those couples which implied a movement of the user in a specific speed range.

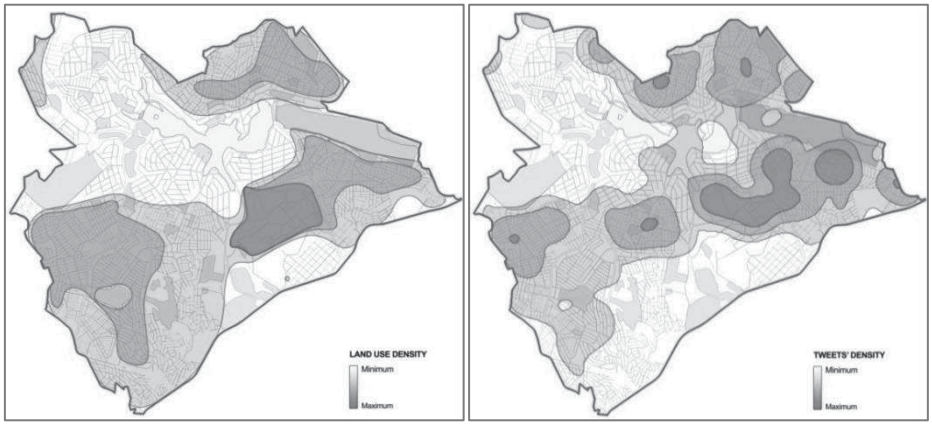

Fig. 2. Quartic Kernel function applied to the traditional land-use map (on the left) and to the footprint map of tweets (on the right).

Then, those couples were used as origin-destination points and overlapped to the actual road network as provided by MapQuest, an OpenStreetMap tool (www.mapquest.com/). Through the shortest path algorithm of the tool itself, a map was created to show the possible paths on the road network that Twitter users could have covered between two subsequent tweets (Figure 3).

Fig. 3. Application of the shortest path algorithm to the couples of subsequent tweets.

Results and discussion

The comparison between traditional map and map derived from social media data was performed on density parameter. Two main comparisons were developed, one for the land use and one for the use of transport infrastructure.

Density of land use

Land-use density map and tweet density map are both generated with the quartic kernel function on the basis of available data. The comparison was performed overlapping and subtracting the two density maps shown in Figure 2, obtaining a further map which illustrates the areas where the two maps provide different information (Figure 4, left). The overlapping of maps provides matching information for the 56% of the area (the white areas in Figure 4, left), while the 42% of area shows light differences between the two maps. A remaining 2% does not match at all (the black areas), so that a deeper investigation is required. The area with more differences of information between the two maps covers the plots of UFMG and sport facilities (Figure 4, centre and right).

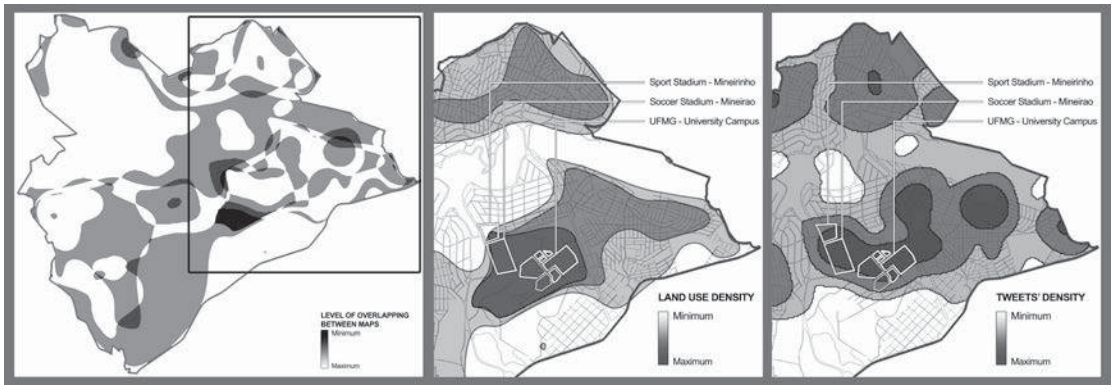

Fig. 4. On the left: overlapping of land-use density and tweet density maps: level of homogeneity of information given by the intersection of areas with the same level of density. On the centre-right: zoom on the area with more heterogeneous information: the official land-use density map (on the centre) and tweet density map (on the right).

Differences are mainly due to the nature of maps: the traditional map illustrates the building or plot density while the tweet maps show the density of (tweeting) people. The grouping mode at the basis of the maps produces different outcomes, which strongly affect both the resulting maps. For instance, the traditional map considers as a high dense area the whole area covered by the university campus, although the South-West zone of the campus is partly un-built. Something similar happens between the soccer stadium and the campus area. This is a green area that, in official maps, appears as a high dense area. The tweet density map shows more details on the use of the area than official maps, highlighting the presence of people in specific areas and offering a more actual view of the use of space. Nevertheless, official land-use map can be more reliable if the planning and decision-making processes should have to consider the possibilities for a future use of the area instead of actual one.

Density of use of transport infrastructures

The density function applied to the official map of transport infrastructures highlights the areas with more vehicles capacity, while the application of shortest path algorithm to sub-sequent tweets provides information on the possible use of road infrastructures. The overlapping of the two information sources shows consistencies when considering the tweeting traffic, which lays

on high-capacity roads, confirming the possibility for many people to pass by those roads. Nevertheless, it shows also divergences. This is the case of the three areas (A, B and C) (Figure 5), in which the high capacity of roads does not correspond to the presence of a large number of people. The roads in A and B are on the North of the lake, while C is a very large roads that goes outside Pampulha. Therefore, some considerations should be done on the use of these high capacity roads.

Fig. 5. Uses of paths according to the subsequent tweets, overlapped on the map of transportation capacity in term of number of vehicles per hour.

If C is a crossing road, it is possible that tweets are not representative of the traffic on that road because not sent during the crossing of the neighbourhood. In that case, the analysis should be widened in order to include all the tweets sent from the surrounding of the area. Furthermore, it should be considered as a possibility that the road in C was not usable in that period. Also in that circumstance, the tweets could not provide usable information. Third option, the high capacity is really over-estimated with respect to the need of people.

Conclusions

The comparative analysis between traditional official maps and maps generated by the analysis of Twitter data confirms the need for a strong filtering of raw data, especially to avoid simple correlations instead of cause-effect relationships (Miller and Goodchild 2015; Masala and Pensa 2016), but it shows also interesting opportunities for integrating and improving the traditional methodologies of planning and decision-making.

In particular, the use of Web social network data provides useful information on the actual use of land. Although tweets are sent by a self-selected population, they represent a mirror of a specific reality.

Finally, data do not provide unique information. This implies a stronger effort from planners and decision-makers in detecting useful information. Thus, data-driven approaches should provide data to the actors involved in the decision-making, allowing the actors to play with data in order to explore different combinations and discover correlations, cause-effects relationships and the best strategies for converting actual dynamics into elements for a well-working spatial system.

References

Bawa-Cavia, Anil. 2010. “Sensing the Urban. Using location-based social network data in urban analysis.” 1st workshop on Pervasive URBan Applications PURBA ‘11. San Francisco.

Chua, Alvin, Ernesto Marcheggiani, Loris Serrvillo, and Andrew Vande Moere. 2014. “FlowSampler: Visual Analysis of Urban Flows in Geolocated Social Media Data.” Edited by Maria Aiello and D McFarland. International Conference on Social Informatics. Barcelona.

García-Palomares, J, C, J Gutiérrez, and C Mínguez. 2015. “Identification of tourist hot spots based on social networks: A comparative analysis of European metropolises using photo-sharing services and GIS.” Applied Geography 63:408-417.

Hahmann, Stefan, Ross, S Purves, and Dirk Burghardt. 2014. “Twitter location (sometimes) matters: Exploring the relationship between georeferenced tweet content and nearby feature classes.” Journal of spatial Information Science (JOSIS) 9:1-36.

Kokalitcheva, Kia. 2014. «How a swarm of data is helping Chicago re-map urban life.» Venture Beat, 14 May. http://venturebeat.com/2014/05/14/how-a-swarm-of-data-is-helping-chicago-re-map-urban-life/ (accessed January 12, 2015).

Kwan, Mei-Po. 2016. «Algorithmic Geographies: Big Data, Algorithmic Uncertainty, and the Production of Geographic Knowledge.» Annals of the American Association of Geographers 106, 2:274-282.

Lanzerotti, Laura, Jeff Bradach, Stephanie Sud, and Henry Barmeier. 2013. “Geek Cities: How Smarter Use of Data and Evidence Can Improve Lives.” The Bridgespan Group, 12 November. http://www.bridgespan.org/Publications-and-Tools/Performance-Measurement/Geek-Cities-Data-Improves-Lives.aspx#.VLZ9BSvF9Ks (accessed January 13, 2015).

Masala, Elena, and Stefano Pensa. 2016. “Detecting spatial features from data-maps: the visual intersection of data as support to decision-making.” RiUS IV: forthcoming.

Miller, Harvey J, and Michael F Goodchild. 2015. “Data-driven geography.” GeoJournal 80, 4:449-461.

Neuhaus, Fabian. 2011. “Twitter Data - Seeking Spatial Pattern.” UrbanTick, 2 March. http://urbantick.blogspot.it/2011/03/twitter-data-seeking-spatial-pattern.html?utm_source=feedburner&utm_medium=feed&utm_campaign=Feed:+urbantick+(urbanTick) (accessed January 09, 2015).

Silverman, B, W. 1986. Density Estimation for Statistics and Data Analysis. New York: Chapman and Hall.

Zachariadis, Vassilis, Camilo Vargas-Ruiz, Joan Serras, e Peter Ferguson. 2015.«Decoding Retail Location: A Primer for the Age of Big Data and Social Media.» CUPUM 2015: The 14th International Conference on Computers in Urban Planning and Urban Management. MIT, Cambridge, MA.